5/22/18

"Coming together is a beginning keeping together is progress working together is success."

-Henry Ford

HW: "Regression" homework section

Test 2 On Wednesday 5/30

AIM: What is Regression?

Warm Up:

2. $\lim\limits_{x \to -3} \dfrac{x^2-9}{2x^2+7x+3}$

3. $\lim\limits_{h \to 0} \dfrac{(2+h)^3-8}{h}$

4. $\lim\limits_{x \to 1} \dfrac{x^8-1}{x^5-x}$

NAME____

A2CC: THE NORMAL DISTRIBUTION -- HOMEWORK 2

1. A population has a mean of $\mu = 24.8$ and a standard deviation of $\sigma = 4.2$. For each of the following data values, calculate the z-value to the nearest hundredth. ~~You do *not* need to read the Normal table.~~

(a) $x_i = 30$ $z = \frac{30-24.8}{4.2} = 1.24$

(b) $x_i = 35$ $z = \frac{35-24.8}{4.2} = 2.43$

(c) $x_i = 19$ $z = -1.38$

(d) $x_i = 15.4$ $z = -2.94$

(e) $x_i = 24.8$ $z = 0$

(f) $x_i = 33.2$ $z = 2.63$

2. A population has a mean of 102.8 and a standard deviation of 15.4. If a data point has a z-value of 1.87 then which of the following is the value of the data point?

(1) 28.8 (3) 131.6

(2) 86.7 (4) 152.3

$z = \frac{x-\mu}{\sigma}$

$1.87 = \frac{x-102.8}{15.4}$

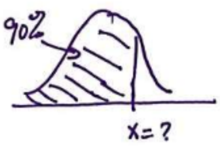
$28.798 = x - 102.8$

$131.598 = x$

3. The weights of four year old boys are normally distributed with a mean of 38 pounds and a standard deviation of 4 pounds. Which of the following weights could represent the 90[th] percentile for the weight of a four year old?

(1) 47 pounds (3) 43 pounds

(2) 45 pounds (4) 41 pounds



$\mu = 38$
$\sigma = 4$

use invnorm

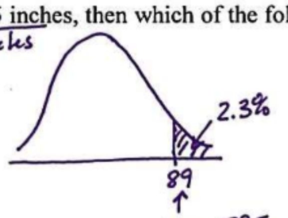
4. The heights of professional basketball players are normally distributed with a standard deviation of 5 inches. If only 2.3% of all pro basketball players have heights above 7 foot 5 inches, then which of the following is the mean height of pro basketball players?

89 inches

(1) 6 feet 5 inches (3) 6 feet 10 inches

(2) 6 feet 2 inches (4) 6 feet 7 inches

$\sigma = 5$
$\mu = ?$



$z = 1.995$

$1.995 = \frac{89-\mu}{5}$

$\mu = 79$ inches

$\mu = 45753$ $\sigma = 7890$

5. If the average teacher salary in the United States is $45,753 and salaries are normally distributed with a standard deviation of $7890, what salary would be at the 40th percentile? 95th percentile?

$43,754 $58,731

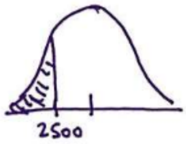
use invnorm for each

6. The average rent for a one bedroom apartment (in the Winter of 2015) in New York City is a whopping $2801 per month with a standard deviation of $920.

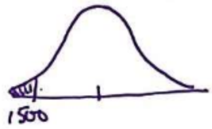
$\mu = 2801$, $\sigma = 920$

(a) If rents are normally distributed, what percent of the apartments will be less than $2,500 per month?

37.2%



(b) If rents are normally distributed, how realistic is it to believe you will be able to rent a one-bedroom in New York City for less than $1,500 per month? Justify your answer.



Only 7.9% of apartments have rents below $1500.

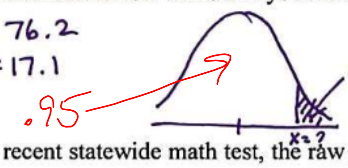
(c) A one-bedroom on the Upper East Side with a doorman and views of Central Park was listed at $5,000 per month. How rare is this? Assume the rents are normally distributed.

Only 0.8% of all apartments have rents at least this high.

(d) Do you think the rents are normally distributed? Keep in mind the normal distribution is symmetric about its mean (looks the same on both sides). If it isn't symmetric, what does it look like? not likely
prob "skewed right"

7. A national math competition advances to the second round only the top 5% of all participants based on scores from a first round exam. Their scores are normally distributed with a mean of 76.2 and a standard deviation of 17.1. What score, to the nearest whole number, would be necessary to make it to the second round? ~~To start, look at the table and see if you can determine the z-value that corresponds to the top 5%.~~

$\mu = 76.2$
$\sigma = 17.1$



use invnorm

$\boxed{104}$

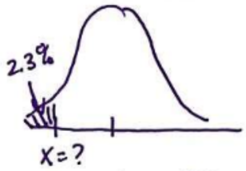
8. On a recent statewide math test, the raw score average was 56 points with a standard deviation of 18. If the scores were normally distributed and 24,000 students took the test, answer the following questions.

$\mu = 56$, $\sigma = 18$

(a) The state would like no more than 550 of the 24,000 students to fail the exam. What percent of the total does the 550 represent? Round to the nearest tenth of a percent.

$\frac{550}{24000} \approx 2.3\%$

(b) What should the raw passing score be set at so that no more than the 550 students fail?



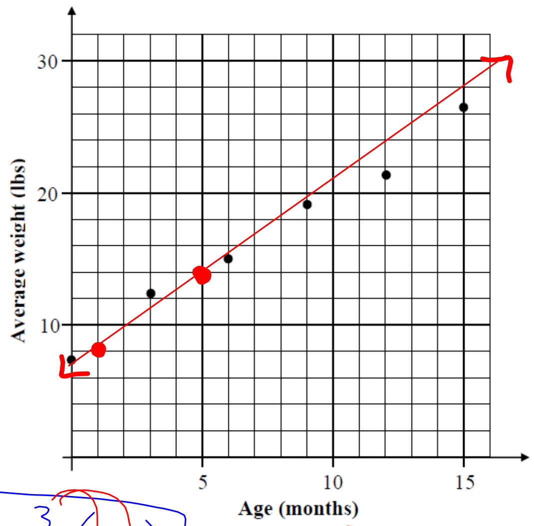
use invnorm

$\boxed{20}$

***Exercise* #1:** A pediatrician would like to determine the relationship between infant female weights versus age. The pediatrician studies 100 newborn girls and finds their average weight at the end of 3 month intervals. The data is shown below and graphed on the scatter plot.

Age (months)	0	3	6	9	12	15
Average Weight (pounds)	7.2	12.2	15.1	19.4	21.5	26.3

(a) Using a ruler, draw a line that you think best fits this data. As a general guideline, try to draw it such that there are as many data points above the line as below it.



(*) Point slope form $y-y_1=m(x-x_1)$

(b) By picking two points that are on the line (not necessarily data points), determine the equation of your best fit line. Round your coefficients to the nearest *tenth*.

(1,8) and (5,14)

Slope $= \frac{14-8}{5-1} = \frac{6}{4} = \frac{3}{2}$

$\boxed{y-8=\frac{3}{2}(x-1)}$ $\Rightarrow$ $y-8=\frac{3}{2}x-\frac{3}{2}$

$+8 \qquad +8$

$y=\frac{3}{2}x+\frac{13}{2}$

$\boxed{y=1.5x+6.5}$

(c) Using the linear regression command on your calculator, find the equation of the best fit line for this data. Round all **linear parameters** to the nearest *tenth*.

```
NORMAL FLOAT AUTO REAL RADIAN MP
LinReg
y=ax+b
a=1.216190476
b=7.828571429
r²=.9905869489
r=.9952823463
```

$y=1.2x+7.8$

(d) Use your calculator to determine the **linear correlation coefficient.** Round to the nearest *thousandth*. How can you interpret this value in terms of the variation in weight due to age?

$r=.995$

close to 1 therefore the line of best fit is a good predictor

(*) Therefore age and weight are highly correlated

***Exercise* #2:** Using the equation that your calculator produced in Exercise #1, predict the weight of a baby girl after 10 months. Round your answer to the nearest tenth of a pound.

$y = 1.2x + 7.8$

$y = 1.2(10) + 7.8$

$\boxed{y = 19.8 \text{ pounds}}$

The use of a model to predict outputs when the input is within the range of the known data is called **interpolation**. Interpolation tends to be fairly accurate.

***Exercise* #3:** Using the equation that your calculator produced in Exercise #1, predict the weight of a baby girl after 2 years. Round your answer to the nearest tenth of a pound.

24 months

$y = 1.2(24) + 7.8$

$y = \boxed{36.6 \text{ pounds}}$

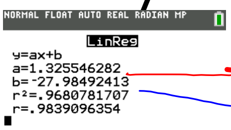
The use of a model to predict outputs when the input is outside of the range of the known input data is called **extrapolation**. Models are most helpful when they can be used to extrapolate, but tend to be less accurate.

***Exercise* #4:** Biologists are trying to create a least-squares regression equation (another name for best fit line) relating the length of steelhead salmon to their weight. Seven salmon were measured and weighed with the data given below.

Length (inches)	22	24	28	34	39	42	48
Weight (pounds)	3.43	4.46	7.08	14.21	22.19	31.22	35.67

(a) Determine the least-squares regression equation, in the form $y = ax + b$, for this data. Round all coefficients to the nearest *hundredth*.

$y = 1.33x - 27.98$



(b) Using your equation from part (a), determine the expected weight of a salmon that is 30 inches long.

$y = 1.33(30) - 27.98$

$= \boxed{11.92 \text{ pounds}}$

(c) Using your equation from part (a), determine the expected weight of a salmon that is 52 inches long.

$y = 1.33(52) - 27.98$

$= \boxed{41.18 \text{ pounds}}$

(d) In which part, (b) or (c), did you use interpolation and in which part did you use extrapolation? Explain.

(b) is interpolation

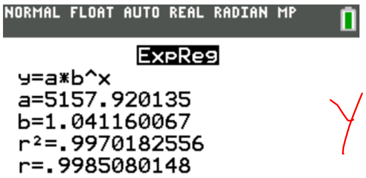
(c) is extrapolation

Just as we fit data with a linear model we can also fit with all sorts of other mathematical models, depending on the context of the situation. In this lesson we will examine **exponential regression** and ~~**sinusoidal regression**~~. You could be asked to run a quadratic regression, logarithmic regression, power regression, …. The process is similar for each and all are found in Stat Calc menu. Exponential regression is review from Common Core Algebra I, so we will start with that.

***Exercise* #5:** The population of Jamestown has been recorded for selected years since 2000. The table below gives these populations.

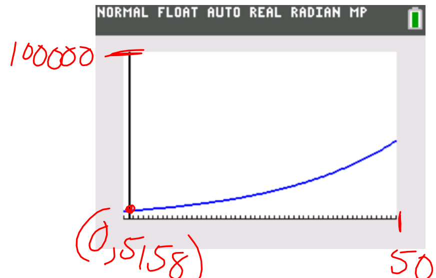
		$x=2$	4	5	7	9
L1	Year	2002	2004	2005	2007	2009
L2	Population	5564	6121	6300	6812	7422

(a) Using your calculator, determine a best fit exponential equation, of the form $y = a \cdot b^x$, where x represents the number of years since 2000 and y represents the population. Round a to the nearest integer and b to the nearest *thousandth*.



$$Y = 5158(1.041)^x$$

(b) Sketch a graph of the exponential function for the years 2000 to 2050. Label your window and your y-intercept.



(c) By what percent does your exponential model predict ~~the population is increasing~~ per year? Explain.

growth rate

$(1.041)^x$

4.1%

(d) Algebraically determine the number of years, to the nearest year, for the population to reach 20 thousand.

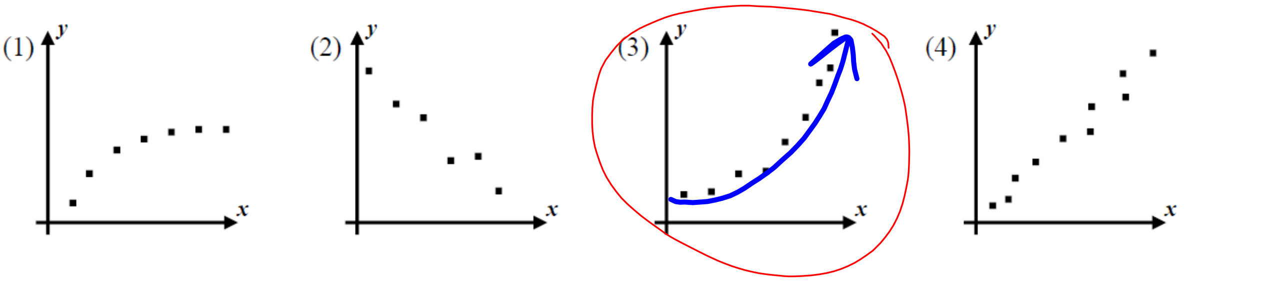
$$\frac{20000}{5158} = \frac{5158(1.041)^x}{5158}$$

$$\frac{20000}{5158} = 1.041^x$$

$$x = \log_{1.041} \frac{20000}{5158}$$

$$\boxed{x \approx 34 \text{ years}}$$

***Exercise* #6:** Which of the following scatter plots would be best fit with an exponential equation?



Sinusoidal, or trigonometric, regression is much more complicated than either linear or exponential. It should be used in situations that appear **periodic** in nature.

***Exercise* #7:** The temperature of a chemical reaction changes during the reaction. The temperature was measured every two minutes and the data is shown in the table below.

Time (min)	0	2	4	6	8	10	12	14	16	18	20
Temp $(^\circ C)$	35.7	38.9	41.6	42.3	40.8	38.4	36.1	34.2	35.9	39.1	41

(a) Why does it seem like this data might be periodic? Create a quick scatter plot using your calculator to verify.

(b) Use your calculator to do a sine regression in the form $y = a\sin(bx+c)+d$. Round all parameters to the nearest tenth. Graph along with your data to informally assess the fit of the curve. When prompted use 16 iterations always.

(c) According to this model, what is the range in temperatures the chemical reaction will include?

(d) According to this model, what is the time it takes for the reaction to complete one full cycle?

***Exercise* #8:** The maximum amount of daylight that hits a spot on Earth is a function of the day of the year. Taking $x = 0$ to be January 1st, daylight, in hours, was measured for 12 different days. The measurement was the number of possible hours of sun from sunrise to sunset.

Day	0	34	68	98	118	134	171	203	274	321	346
Daylight Hours	9.0	9.9	11.5	13.1	14.0	14.6	15.2	14.8	13.1	11.5	9.5

(a) What is the natural period of this data set?

(b) Use your calculator with the period from (a) to find an equation of the form $y = a\sin(bx+c)+d$ that fits this data, then examine the graph of the equation on the scatter plot. How good is the fit?

(c) What is the maximum amount of daylight hours predicted by the model? Show your calculation.